NEAR-OPTIMAL TIME-SPACE TRADEOFF FOR ELEMENT DISTINCTNESS

Andrew Chi-Chih Yao

CS-TR-139-88

March 1988

Near-Optimal Time-Space Tradeoff for Element Distinctness

Andrew Chi-Chih Yao

Department of Computer Science
Princeton University
Princeton, New Jersey 08544

Abstract

It was conjectured in Borodin et al. [*J. Comput. System Sci.* **22** (1981), pp. 351-364] that to solve the element distinctness problem requires $TS = \Omega(n^2)$ on a comparison-based branching program using space S and time T, which, if true, would be close to optimal since $TS = O(n^2 \log n)$ is achievable. Recently, Borodin et al. [*SIAM J. on Comput.* **16** (1987), pp. 97-99] showed that $TS = \Omega(n^{3/2}(\log n)^{1/2})$. In this paper, we will show a near-optimal tradeoff $TS = \Omega(n^{2-\epsilon(n)})$, where $\epsilon(n) = O(1/(\log n)^{1/2})$.

1 Introduction

In Cobham's classic paper [C], time-space tradeoffs were established for one-tape Turing machines. In recent years, a number of time-space tradeoff results have been obtained for various computational models, such as Boolean and arithmetic circuits (Tompa [To]), a general sequential computing model (Borodin and Cook [BC]), multihead Turing machines (Duris and Galil [DG], Karchmer [K]), comparison-based branching programs (Borodin et al. [BFKLT], Yao [Y], Borodin et al. [BFMUW], Johnson [J], Karchmer [K]), and VLSI models (Thompson [Th], see Ullman [U] for a review of results). In this paper, we will establish a tradeoff result in the comparison-based branching program model, proving in weaker form an interesting conjecture of Borodin et al. [BFKLT].

Borodin et al. [BFKLT] proved a tradeoff $TS = \Omega(n^2)$ for sorting n numbers on a comparison-based branching program, but were not able to establish a similar tradeoff for any decision problem in their model. They conjecture, however, that the tradeoff $TS = \Omega(n^2)$ is also true for the element distinctness problem. This, if proved, would be close to the best possible, since an upper bound $TS = O(n^2 \log n)$ is achievable for sorting, and hence for the element distinctness problem. Recently, Borodin et al. [BFMUW] gave a partial resolution to the above conjecture, showing in the same model that $TS = \Omega(n^{3/2}(\log n)^{1/2})$. In this paper, we will prove that $TS = \Omega(n^{2-\epsilon(n)})$, where $\epsilon(n) = O(1/(\log n)^{1/2})$. As mentioned earlier, such a tradeoff is nearly the best possible.

Let $x_1, x_2, \ldots, x_n$ be n elements chosen from a linearly ordered set $(D, \leq)$. The element distinctness problem (on n elements) is to decide whether all x_i are distinct. Following [BFMUW], a *comparison branching program* A is a labeled directed acyclic graph with a distinguished nonsink node, called the *source*. Each nonsink node is labeled by a comparison $x_i : x_j$ with $i \neq j$, and has three outgoing edges, labeled by $<$, $=$, $>$, respectively. The sinks are labeled by either "accept" or "reject". An input $\tilde{x} = (x_1, x_2, \ldots, x_n) \in D^n$ starts at the source and traverses A, making comparisons and branching according to the outcomes, until a sink is reached. The input is accepted if and only if it reaches a sink with an "accept" label. The *capacity* of A is the base-2 logarithm of the number of nodes. The *length* of A, or the *time* T used by A, is the length of the longest path starting with source. We say that A is an *algorithm* for *the element distinctness problem*, if $\tilde{x}$ is accepted when and only when all x_i are distinct. Let $\mathcal{A}_n$ denote the set of all algorithms for the element distinctness problem. We now state our main result.

Theorem 1 Any $A \in \mathcal{A}_n$ with capacity S and time T must satisfy $TS = \Omega(n^{2-\epsilon(n)})$ for large n, where $\epsilon(n) = 5/(\ln n)^{1/2}$.

2 Preliminaries

2.1 Overview

As discussed in [BFKLT], by a result of Nick Pippenger, we can assume without loss of generality that A is *leveled*, i.e. each node is assigned a nonnegative integral level number, and each edge goes from a node at level i to a node at level $i+1$; the source node is the only node at level 0, and all sinks are at level T. From now on, all branching programs will mean leveled comparison branching programs.

We review the ideas involved in the proof of $TS = \Omega(n^{3/2}(\log n)^{1/2})$ in [BFMUW]. (Some of these ideas originated in [BFKLT].) Let $A \in \mathcal{A}_n$. For any input $\tilde{x} = (x_1, x_2, \ldots, x_n)$ with distinct x_i, the sequence of comparisons made by A must include all the "adjacent" ones, i.e. comparisons of the form $x_{i_j} : x_{i_{j+1}}$ if the input satisfies $x_{i_1} < x_{i_2} < \cdots < x_{i_n}$; otherwise we could have two identical x_r. The idea is to show that any branching program of length less than or equal to n_0, where $n_0 = (nS/(16e))^{1/2}$, can make more than S adjacent comparisons only for a very small fraction of the $n!$ possible linear orderings in the input. Thus, if we divide A into consecutive blocks of n_0 levels each, there must be at least $(n-1)/S$ such blocks in order to perform the needed $n-1$ adjacent comparisons for all linear orderings. This proves $T \geq n_0(n-1)/S$, and hence $TS = \Omega(n^{3/2}S^{1/2}) = \Omega(n^{3/2}(\log n)^{1/2})$ as $S = \Omega(\log n)$.

To prove Theorem 1, we will adopt the same general approach. We will show that any branching program of length less than or equal to n_1, where $n_1 = n^{1-\epsilon(n)}$, can make more than $S \cdot n^{\epsilon(n)}$ adjacent comparisons only for a very small fraction of the possible linear orderings. The asserted tradeoff then follows the same line reasoning as before.

2.2 Terminology

Let $W = \{w_1, w_2, \ldots, w_n\}$ be any nonempty finite set. A *linear ordering on* W is a sequence $\sigma = \langle w_{i_1} < w_{i_2} < \cdots < w_{i_n} \rangle$, in which each element of W appears exactly once. Let $\Gamma(W)$ denote the set of all linear orderings on W.

Let $P = (<_P, W)$ be a partial order on W. A linear ordering ρ is said to be *consistent with* P, if $w <_P w'$ implies $w < w'$ in ρ. Let $\Delta(P)$ denote the set of all linear orderings on W consistent with P.

Suppose that $W' \subseteq W$ and $\sigma = \langle w_{r_1} < w_{r_2} < \cdots < w_{r_\ell} \rangle \in \Gamma(W')$. A comparison $w_i < w_j$ is *adjacent in* σ, if w_i, w_j are adjacent in the linear ordering σ, i.e. there exists an m such that $i = r_m$ and $j = r_{m+1}$. For any $\rho \in \Gamma(W)$, let $\rho|_{W'}$ denote the $\sigma \in \Gamma(W')$ obtained from the restriction of ρ to W'. For example, if $\rho = \langle w_3 < w_2 < w_5 < w_1 < w_4 \rangle$ and $W' = \{w_2, w_4, w_5\}$, then $\rho|_{W'} = \langle w_2 < w_5 < w_4 \rangle$.

We will use the symbol X to denote exclusively the set $\{x_1, x_2, \ldots, x_n\}$ of the n input numbers. Let C be a sequence of inequalities $(x_{i_1} < x_{j_1}, x_{i_2} < x_{j_2}, \ldots, x_{i_\ell} < x_{j_\ell})$. Its length ℓ is denoted by $|C|$, and *support*(C) is the set of all x_r involved in any comparisons in C, i.e. those x_r with $r = i_s$ or $r = j_{s'}$ for some s, s'. We say that C is *nontrivial*, if there is a linear ordering $\langle x_{r_1} < x_{r_2} < \cdots < x_{r_n} \rangle$ on X such that all inequalities in C are true. For any nontrivial C, its transitive closure defines a partial order P_C on the set support(C). Now, for each C satisfying $2|C| \leq n$, let $V_C \subseteq X$ be a set disjoint from support(C) with $|V_C| = 2|C| - |\text{support}(C)|$, and define *support*$'(C)$= support$(C) \cup V_C$. Let P'_C be the partial order P_C regarded as a partial order on support$'(C)$. In particular, if all the pairwise comparisons in C are disjoint, then $V = \emptyset$ and $P'_C = P_C$.

Let $0 \leq r_1 < r_2 \leq \ell$ be two integers. For any $\sigma \in \Delta(P'_C)$, let $Z(C, r_1, r_2, \sigma)$ be the number of pairs (i_s, j_{s+1}), $r_1 < s \leq r_2$, such that the comparisons $x_{i_s} < x_{j_s}$ are adjacent in σ.

2.3 Main Lemma

Let A be any branching program of variables $x_1, x_2, \ldots, x_n$. For any node u and positive integer s, let $A[u, s]$ denote the sub-branching program of A of length s and rooted at node u. A *path* δ in A is a sequence $u_1, e_1, u_2, e_2, \ldots, u_s, e_s$, where each e_r is an edge from node u_r to u_{r+1}; s is the *length* of δ. Let C_δ denote the sequence of comparison results "$x_i < x_j$" and "$x_r = x_s$" obtained along the path δ ("$x_i > x_j$" will be written as "$x_j < x_i$"). We are only interested in paths δ that (a) contain no edge labeled by "=", and (b) have nontrivial C_δ. From now on, when we speak of a *path* δ, we require that the above two conditions be satisfied. Let Λ_A be the set of all paths of length T that begin with the source, where T is the length of A. For any linear ordering $\rho \in \Gamma(X)$, let $\tilde{x}_\rho$ denote an input $(x_1, x_2, \ldots, x_n)$ that satisfies all the inequalities in ρ. Let $\xi_{A,\rho} \in \Lambda_A$ be the path traversed by input $\tilde{x}_\rho$.

Let $N_0 = 10^8$, $n \geq N_0$, $S > 0$, $t = \lceil e^{(\ln n)^{1/2}} \rceil$, and $k_0 = \lfloor \log_t(n/4) \rfloor$. Then $t, k_0 \geq 4$. For any integer $k > 0$, let $m_k = 2^{4k+16} t S$, and $q_k = (4t)^k 2^{-10S}$.

Main Lemma Let $1 \leq k \leq k_0$. Suppose A is a branching program of length t^k and capacity S. Take a random ρ, uniformly chosen from $\Gamma(X)$, then $\Pr\{Z(C_\delta, 0, t^k, \rho|_{X'}) \geq m_k\} \leq q_k$, where $\delta = \xi_{A,\rho}$ and $X' = \text{support}'(C_\delta)$.

Corollary Suppose that A is a branching program of length $\leq t^{k_0}$ and capacity S. Take a random ρ, uniformly chosen from $\Gamma(X)$, and input $\tilde{x}_\rho$ to A, then the probability that A makes at least m_k comparisons adjacent in ρ is $\leq q_k$.

We remark that the Main Lemma is true for any choice of the sets V_C. However, the choice must be made before taking the random ρ. The corollary follows from the Main Lemma, since the introduction of additional elements x_i into support$'(C_\delta)$ will not increase the number of adjacent comparisons. It is this corollary that we will use later in the proof of Theorem 1. Before doing that, we need to prove the Main Lemma. We first derive an auxiliary lemma in the next subsection. This will be used in Section 3 to prove the Main Lemma. The proof of Theorem 1 will then be given in Section 4.

2.4 An Auxiliary Lemma

Let $X' \subseteq X$ be nonempty, and $\sigma \in \Gamma(X')$. Suppose that C is a nontrivial sequence of comparisons $x_i < x_j$ with $x_i, x_j \in X'$, with exactly ℓ of them being adjacent in σ. Let A be a branching program of length T with $n \geq 2T + |X'|$. For any path $\delta \in \Lambda_A$, let $W_\delta \subseteq X$ be such that W_δ is disjoint from $X' \cup \text{support}(C_\delta)$, and $|W_\delta| = 2T + |X'| - |X' \cup \text{support}(C_\delta)|$. Let $W'_\delta = W_\delta \cup X' \cup \text{support}(C_\delta)$. Clearly, $|W'_\delta| = 2T + |X'|$.

Now, take a random $\rho \in L(\sigma)$, uniformly chosen, and let $f(\sigma, C, A, m)$ be the probability that the number of comparisons of C adjacent in $\rho|_{W'_\delta}$ is greater than or equal to m, where $\delta = \xi_{A,\rho}$.

Lemma 1 Suppose $n \geq 2T + |X'|$. Then $f(\sigma, C, A, m) \leq \binom{\ell}{m} \cdot \left(\frac{|X'|}{2T}\right)^m$.

Proof Without loss of generality, assume that $X' = \{x_1, x_2, \ldots, x_a\}$, where $a = |X'|$. Let C' be the set of the comparisons of C that are adjacent in σ, say, $C' = \{x_{i_1} < x_{i_1+1}, x_{i_2} < x_{i_2+1}, \ldots, x_{i_\ell} < x_{i_\ell+1}\}$. We can also assume that $\ell > 0$ and $|X| > m > 0$; otherwise the lemma is trivially true.

We express f in terms of a stochastic process. Take a random $\rho \in L(\sigma)$, and traverse the path $\xi_{A,\rho}$. Let us keep a sorted list W'; initially, W' is the sorted version of X'. When we encounter a new node u with a comparison $x_r : x_s$, we insert the elements in $\{x_r, x_s\} - W'$, if any, one at a time into the ordered list W'. Note that each new element, when added to an ordered list of c elements, will be equally likely in any of the $c+1$ ranks. When we reach the leaf, we add the $2T + |X'| - |W'|$ new elements of W_δ, one at a time, into W'. Again, each new element is eqally likely to occupy any of the ranks currently possible in W'. The quantity $f(\sigma, C, A, m)$ can thus be calculated as follows: We start with an ordered list of $|X'|$ items with ℓ of the intervals (between the i_j-th and the i_{j+1}-th items for $1 \leq j \leq \ell$) marked; then we sequentially insert new items into the list, each time the new item is equally likely to be inserted into any of the existing intervals; $f(\sigma, C, A, m)$ is the probability that, after $2T$ insertions, at least m of the original ℓ marked intervals remain intact (no item has been inserted into these intervals).

We will obtain an upper bound on f. (Essentially, this is now reduced to a calculation which was done in [BFMUW].) Let us describe the above stochastic process using a sequence of $2T$ integers $j_1, j_2, \ldots, j_{2T}$, where $1 \leq j_r \leq |X'| + r$ is the rank of the r-th inserted item when it is being inserted. Thus, there are in all $\prod_{1 \leq r \leq 2T}(|X'| + r)$ configurations. To specify a configuration for which at least m marked intervals remain intact, we first specify m such intervals, and then specify the ranks of the inserted items by integers $j_1, j_2, \ldots, j_{2T}$, where $1 \leq j_r \leq |X'| + r - m$.

The total number of such configurations is thus at most $\binom{\ell}{m}\prod_{1\le r\le 2T}(|X'|+r-m)$. It follows that

$$
\begin{aligned}
f(\sigma,C,A,m) &\le \binom{\ell}{m}\frac{\prod_{1\le r\le 2T}(|X'|+r-m)}{\prod_{1\le r\le 2T}(|X'|+r)}\\
&= \binom{\ell}{m}\frac{|X'|!}{(|X'|+2T)!}\frac{(|X'|+2T-m)!}{(|X'|-m)!}\\
&= \binom{\ell}{m}\frac{|X'|\cdot(|X'|-1)(|X'|-m+1)}{(|X'|+2T)(|X'|+2T-1)\cdots(|X'|+2T-m+1)}\\
&\le \binom{\ell}{m}\cdot\left(\frac{|X'|}{2T}\right)^m.\square
\end{aligned}
$$

3 Proof of the Main Lemma

We will prove the Main Lemma by induction on $k\ge 1$. For $k=1$, we have $m_k>t^k$. Since A can make only t^k comparisons, $M_A=\emptyset$, and the Main Lemma is true in this case.

We now assume that $1<k\le k_0$, and that the Main Lemma has been proved for all values less than k.

Assume that the choice of V_C has been made. Let $\rho\in\Gamma(X)$. Write $\delta=\xi_{A,\rho}$ and $X'=\text{support}'(C_\delta)$. If $Z(C_\delta,0,t^k,\rho|_{X'})\ge m_k$, then there is a $1\le d_\rho\le t$, such that $Z(C_\delta,(d_\rho-1)t^{k-1},d_\rho t^{k-1},\rho|_{X'})\ge m_k/t$. Thus,

$$\Pr\{Z(C_\delta,0,t^k,\rho|_{X'})\ge m_k\}\le\sum_{1\le d\le t}\Pr\{Z(C_\delta,(d-1)t^{k-1},dt^{k-1},\rho|_{X'})\ge m_k/t\}. \tag{1}$$

We will show that, for each $1\le d\le t$,

$$\Pr\{Z(C_\delta,(d-1)t^{k-1},dt^{k-1},\rho|_{X'})\ge m_k/t\}\le 2q_{k-1}. \tag{2}$$

This will complete the inductive proof of the Main Lemma, as it follows from (1) and (2) that

$$
\begin{aligned}
\Pr\{Z(C_\delta,0,t^k,\rho|_{X'})\ge m_k\} &\le 2tq_{k-1}\\
&\le q_k.
\end{aligned}
$$

Although we have made the choice of V_C, which is needed to define the function Z above, we observe that the values of $\Pr\{Z(C_\delta,0,t^k,\rho|_{X'})\ge m_k\}$ and $\Pr\{Z(C_\delta,(d-1)t^{k-1},dt^{k-1},\rho|_{X'})\ge m_k/t\}$ are in fact independent of the choice of V_C, as ρ is uniformly chosen from $\Gamma(X)$. We will now evaluate $\Pr\{Z(C_\delta,(d-1)t^{k-1},dt^{k-1},\rho|_{X'})\ge m_k/t\}$ with a special new choice of V_{C_δ} to be described below.

Fix $1 \le d \le t$. Let $v_1, v_2, \ldots, v_r$ be the nodes of A at level $(d-1)t^{k-1}$. For each i, let B_i be the set of paths β of length t^{k-1} starting at node v_i. For each $\beta \in B_i$, let us choose a subset $X_\beta \subseteq X$ such that (a) $X_\beta \cap \text{support}(C_\beta) = \emptyset$, and (b) $|X_\beta| + |\text{support}(C_\beta)| = 2t^{k-1}$. Let $X'_\beta = X_\beta \cup \text{support}(C_\beta)$, and Q_β be the partial order on X'_β generated by the inequalities in C_β. Let $\Psi_i = \cup_{\beta \in B_i} \Delta(Q_\beta)$.

For any $\sigma \in \Gamma(W)$, where $W \subseteq X$, let $L(\sigma)$ denote the set of all linear orderings $\rho \in \Gamma(X)$ that are consistent with σ.

Fact 1 Let $\beta, \beta' \in B_i$. If $\sigma \in \Delta(Q_\beta) \cap \Delta(Q_{\beta'})$, then $\beta = \beta'$.

Fact 2 The family $L(\sigma)$, $\sigma \in \Psi_i$, form a partition of the set $\Gamma(X)$.

Fact 1 is true because any two distinct β, β' must have a common node at which the comparison $x_r : x_s$ made gives opposite outcomes. We can thus write $\beta(i, \sigma)$ for the unique β for which $\sigma \in \Delta(Q_\beta)$. To prove Fact 2, first we observe that every $\tilde{x}_\rho$ starting at v_i will follow some path β. This shows that the union of $L(\sigma)$, $\sigma \in \Psi_i$, contains $\Gamma(X)$. It remains to prove that, if $\sigma \ne \sigma'$, then $L(\sigma) \cap L(\sigma') = \emptyset$. This is clearly true when $\sigma, \sigma' \in \Delta(Q_\beta)$ for some common β. In the other case, $\sigma \in Q_\beta$ and $\sigma' \in Q_{\beta'}$ with $\beta \ne \beta'$. Any $\rho \in L(\sigma)$ and $\rho' \in L(\sigma')$ must be different, since $\tilde{x}_\rho$, $\tilde{x}_{\rho'}$ follow two different paths β, β'. This proves Fact 2.

For the discussion to follow, we will use the convention that a branching program of length 0 is a *null branching program*, denoted by Φ. We agree that the expressions $M\Phi$, ΦM both stand for M, where M is any branching program. A path of length 0 is the *null path*, denoted by ψ. A sequence of comparison inequalities is the *null sequence*, denoted by κ. We define C_ψ to be κ. Define support(κ)$= \emptyset$, and $\Lambda_\Phi = \{\psi\}$. For any $\rho \in \Gamma(X)$, let $\xi_{\Phi,\rho} = \psi$. The introduction of these notations is mainly for convenience, so as to avoid the necessity of discussing degenerate cases in the discussions to come.

Let $A' = A[\text{root}\ , (d-1)t^{k-1}]$, and $A_\beta = A[u_\beta, t^k - t^{k-1}]$, where u_β is the node reached by the last edge of the path β. Let $A'A_\beta$ denote the branching program one obtains by attaching a copy of A_β to each leaf of A'. The length of $A'A_\beta$ is clearly $t^k - t^{k-1}$. We remark that, if $d = 1$, then $A' = \Phi$; if $d = t$, then $A_\beta = \Phi$ for all β.

Let $\eta = (i, \alpha, \beta, \gamma)$ be any quadruple, where $1 \le i \le r$, $\alpha \in \Lambda_{A'}$, $\beta \in B_i$, and $\gamma \in \Lambda_{A_\beta}$. Define $U_\eta = \text{support}(C_\alpha) \cup \text{support}(C_\gamma) \cup X'_\beta \cup W_{\alpha,\gamma}$, where $W_{\alpha,\gamma} \subseteq X$ satisfies the conditions that it is disjoint from $\text{support}(C_\alpha) \cup \text{support}(C_\gamma) \cup X'_\beta$ and has cardinality $2t^k - |\text{support}(C_\alpha) \cup \text{support}(C_\gamma) \cup X'_\beta|$ but arbitrary otherwise. Thus, $|U_\eta| = 2t^k$. Note that the last edge of the path α does not have to end in the node v_i. For any $\rho \in \Gamma(X)$, define $Y(\rho, \eta) = 1$, if the number of comparisons of C_β adjacent in $\rho|_{U_\eta}$ is at least m_k/t, and 0 otherwise.

Let $\delta \in \Lambda_A$. We will describe how to choose V_{C_δ}. We can uniquely write $\delta = \alpha\beta\gamma$, where for some i, α is a path in A', $\beta \in B_i$, and γ is a path in A_β. Let $\eta(\rho) = (i, \alpha, \beta, \gamma)$. Define

$V_{C_\delta} = (X_\beta \cup W_{\alpha,\gamma}) - \text{support}(C_\alpha) - \text{support}(C_\gamma)$. Then $\text{support}'(C_\delta) = V_{C_\delta} \cup \text{support}(C_\delta) = X'_\beta \cup \text{support}(C_\alpha) \cup \text{support}(C_\gamma) \cup W_{\alpha,\gamma}$. Thus, $\text{support}'(C_\delta) = U_{\eta(\rho)}$. It follows that $Y(\rho, \eta(\rho)) = 1$ if and only if $Z(C_\delta, (d-1)t^{k-1}, dt^{k-1}, \rho|_{X'}) \geq m_k/t$.

For $1 \leq i \leq r$, let R_i be the set of $\rho \in \Gamma(X)$ such that input $\tilde{x}_\rho$ will reach v_i in A. Let $\Psi_{i,1}$ be the set of $\sigma \in \Psi_i$ such that $C_{\beta(i,\sigma)}$ contains at least m_{k-1} comparisons adjacent in σ. Let $\Psi_{i,2}$ be the set of $\sigma \in \Psi_i$ such that $|L(\sigma) \cap R_i| \leq |L(\sigma)| q_{k-1}/(10 \cdot 2^S)$. Let $\Psi_{i,0} = \Psi_i - \Psi_{i,1} - \Psi_{i,2}$.

Let p_i be the probability that $\xi_{A,\rho}$ will reach v_i for a random ρ, i.e. $p_i = |R_i|/n!$. Let I be the set of i with $p_i > 0$. For any $i \in I$ and $\sigma \in \Psi_i$, let $p_{i,\sigma} = |L(\sigma) \cap R_i|/|R_i|$, i.e. the probability that, given that v_i is reached by $\xi_{A,\rho}$, ρ will be consistent with σ. Let $\Psi'_i = \{\sigma \mid p_{i,\sigma} > 0\}$. Define $\Psi'_{i,j} = \Psi_{i,j} \cap \Psi'_i$ for $j \in \{0, 1, 2\}$. With the above choice of V_{C_δ} for defining Z, we have

$$\begin{aligned} \Pr\{Z(C_\delta, (d-1)t^{k-1}, dt^{k-1}, \rho|_{X'}) &\geq m_k/t\} \\ &= \sum_{i \in I} \sum_{\sigma \in \Psi'_i} p_i p_{i,\sigma} \Big\{ \frac{1}{|L(\sigma) \cap R_i|} \sum_{\rho \in L(\sigma) \cap R_i} Y(\rho, \eta(\rho)) \Big\}. \quad (3) \end{aligned}$$

We need three facts. Let $i \in I$.

Fact 3 $\sum_{\sigma \in \Psi'_{i,1}} p_{i,\sigma} \leq q_{k-1}$.

Fact 4 $\sum_{\sigma \in \Psi'_{i,2}} p_i p_{i,\sigma} \leq q_{k-1}/(10 \cdot 2^S)$.

Fact 5 For each $\sigma \in \Psi'_{i,0}$ with $L(\sigma) \cap R_i \neq \emptyset$,

$$\frac{1}{|L(\sigma) \cap R_i|} \sum_{\rho \in L(\sigma) \cap R_i} Y(\rho, \eta(\rho)) \leq q_{k-1}/10.$$

If we apply the induction hypothesis to $A[v_i, t^{k-1}]$, we get Fact 3. We obtain Fact 4 from the following derevation, using Fact 2 in the last step,

$$\begin{aligned} \sum_{\sigma \in \Psi'_{i,2}} p_i p_{i,\sigma} &= \sum_{\sigma \in \Psi'_{i,2}} \frac{|R_i|}{n!} \frac{|L(\sigma) \cap R_i|}{|R_i|} \\ &= \frac{1}{n!} \sum_{\sigma \in \Psi'_{i,2}} |L(\sigma)| \frac{|L(\sigma) \cap R_i|}{|L(\sigma)|} \\ &\leq \frac{1}{n!} \frac{1}{10 \cdot 2^S} q_{k-1} \sum_{\sigma \in \Psi'_{i,2}} |L(\sigma)| \\ &\leq \frac{1}{10 \cdot 2^S} q_{k-1}. \end{aligned}$$

We now prove Fact 5. Let $i \in I$ and $\sigma \in \Psi'_{i,0}$. Take a random ρ, uniformly chosen from $L(\sigma)$. Write $A'' = A_{\beta(i,\sigma)}$ and $\zeta(\rho) = (i, \xi_{A',\rho}, \beta(i,\sigma), \xi_{A'',\rho})$. Then

$$\begin{aligned}
\frac{1}{|L(\sigma)\cap R_i|}\sum_{\rho\in L(\sigma)\cap R_i} Y(\rho,\eta(\rho)) &= \frac{1}{|L(\sigma)\cap R_i|}\sum_{\rho\in L(\sigma)\cap R_i} Y(\rho,\zeta(\rho)) \\
&\leq \frac{1}{|L(\sigma)\cap R_i|}\sum_{\rho\in L(\sigma)} Y(\rho,\zeta(\rho)) \\
&= \frac{|L(\sigma)|}{|L(\sigma)\cap R_i|}\frac{1}{|L(\sigma)|}\sum_{\rho\in L(\sigma)} Y(\rho,\zeta(\rho)) \\
&\leq \frac{10\cdot 2^S}{q_{k-1}}\frac{1}{|L(\sigma)|}\sum_{\rho\in L(\sigma)} Y(\rho,\zeta(\rho)) \qquad (4)
\end{aligned}$$

It is clear that

$$\frac{1}{|L(\sigma)|}\sum_{\rho\in L(\sigma)} Y(\rho,\zeta(\rho)) = f(\sigma, C_{\beta(i,\sigma)}, A'A_{\beta(i,\sigma)}, m_k/t). \qquad (5)$$

Let ℓ_σ denote the number of comparisons of $C_{\beta(i,\sigma)}$ adjacent in σ. Then $\ell_\sigma \leq m_{k-1}$ as $\sigma \in \Psi'_{i,0}$. Using (5) and Lemma 1 (noting that $n \geq 2t^k$), we have

$$\begin{aligned}
\frac{1}{|L(\sigma)|}\sum_{\rho\in L(\sigma)} Y(\rho,\zeta(\rho)) &\leq \binom{\ell_\sigma}{\lceil m_k/t\rceil}\left(\frac{2t^{k-1}}{2(t^k - t^{k-1})}\right)^{\lceil m_k/t\rceil} \\
&\leq \binom{m_{k-1}}{\lceil m_k/t\rceil}(t-1)^{-\lceil m_k/t\rceil} \\
&\leq \left(\frac{em_{k-1}}{(t-1)\lceil m_k/t\rceil}\right)^{\lceil m_k/t\rceil} \\
&\leq \left(\frac{et}{10(t-1)}\right)^{m_k/t} \\
&\leq \frac{1}{2^{1000S}} \qquad (6)
\end{aligned}$$

It follows from (4) and (6) that

$$\begin{aligned}
\frac{1}{|L(\sigma)\cap R_i|}\sum_{\rho\in L(\sigma)\cap R_i} Y(\rho,\eta(\rho)) &\leq \frac{10\cdot 2^S}{q_{k-1}}\left(\frac{1}{2}\right)^{1000S} \\
&\leq \frac{1}{10}q_{k-1}.
\end{aligned}$$

This proves Fact 5.

We will now complete the proof of (2). From Facts 3, 4 and 5, we obtain from (3) that

$$\Pr\{Z(C_\delta, (d-1)t^{k-1}, dt^{k-1}, \rho|_{X'}) \geq m_k/t\}$$

$$
\begin{aligned}
&= \sum_{i\in I}\sum_{\sigma\in\Psi'_i} p_i p_{i,\sigma}\Big\{\frac{1}{|L(\sigma)\cap R_i|}\sum_{\rho\in L(\sigma)\cap R_i} Y(\rho,\eta(\rho))\Big\} \\
&\le \sum_{i\in I}\sum_{\sigma\in\Psi'_{i,1}} p_i p_{i,\sigma} + \sum_{i\in I}\sum_{\sigma\in\Psi'_{i,2}} p_i p_{i,\sigma} \\
&\quad + \sum_{i\in I}\sum_{\sigma\in\Psi'_{i,0}} p_i p_{i,\sigma}\frac{1}{|L(\sigma)\cap R_i|}\sum_{\rho\in L(\sigma)\cap R_i} Y(\rho,\eta(\rho)) \\
&\le q_{k-1}\sum_{i\in I} p_i + \sum_{i\in I}\frac{1}{10\cdot 2^S} q_{k-1} + \sum_{i\in I}\sum_{\sigma\in\Psi'_{i,0}} p_i p_{i,\sigma}\frac{1}{10} q_{k-1} \\
&\le q_{k-1} + \frac{1}{10} q_{k-1} + \frac{1}{10} q_{k-1} \\
&\le 2q_{k-1}.
\end{aligned}
$$

We have proved (2). This completes the inductive step in the proof of the Main Lemma.

4 Proof of Theorem 1

Let $n \ge N_0$, where $N_0 = 10^8$. Suppose that $A \in \mathcal{A}_n$ has capacity S and time T. Define t, k_0, m_k, q_k as in Section 2.3. Clearly,

$$T \ge n - 1, \tag{7}$$

and

$$S \ge \lg(n-1). \tag{8}$$

Also, $k_0 \le \ln n/\ln t \le (\ln n)^{1/2}$, and hence

$$
\begin{aligned}
t2^{4k_0} &\le e^{(\ln n)^{1/2} + k_0 4\ln 2} \\
&\le n^{\epsilon(n)}
\end{aligned}
\tag{9}
$$

It follows that $m_{k_0} = 2^{4k_0+16} tS \le 2^{16} n^{\epsilon(n)} S$. If $m_{k_0} \ge n/8$, then $S = \Omega(n^{1-\epsilon(n)})$; hence from (7) we have $TS = \Omega(n^{2-\epsilon(n)})$, which proves the theorem. Thus, we can assume that

$$m_{k_0} < \frac{n}{8}. \tag{10}$$

We will prove that

$$T \ge \frac{t^{k_0} n}{2m_{k_0}}. \tag{11}$$

Suppose not. We will derive a contradiction. For each node v, let F_v denote the branching program $A[v, \min\{t^{k_0}, T - h_v\}]$, where h_v is the level number of v. Let K_v be the set of $\rho \in \Gamma(X)$

such that $C_{\delta(v,\rho)}$ contains at least m_{k_0} comparisons adjacent in ρ, where $\delta(v,\rho) = \xi_{F_v,\rho}$. By the corollary to the Main Lemma, $|K_v| \leq q_{k_0} \cdot n!$.

Thus,

$$|\cup_v K_v| \leq 2^S q_{k_0} \cdot n! \tag{12}$$

Since $t \geq 4$ and $t^{k_0} \leq n/4$, we have $q_{k_0} \leq t^{2k_0} 2^{-10S} \leq n^2 2^{-10S}$. Using (8), we have $2^S q_{k_0} \leq n^2(n-1)^{-9} < 1$. Therefore, (12) implies that there exists a $\rho \notin \cup_v K_v$. Let us input $\tilde{x}_\rho$ to A. The total number of comparisons made by A that are adjacent in ρ is less than $\lceil T/t^{k_o} \rceil \cdot m_{k_0} \leq \lceil n/2m_{k_0} \rceil \cdot m_{k_0} \leq n/2 + m_{k_0}$, which by (10) is less than $n-1$. This is a contradiction. We have proved (11).

It follows from (11) and (9) that

$$\begin{aligned} T &\geq t^{k_0} \frac{n}{2m_{k_0}} \\ &= \Omega\Big(\frac{t^{k_0-1}}{2^{4k_0}} \frac{n}{S}\Big) \\ &= \Omega\Big(\frac{n^2}{S} \frac{1}{t 2^{4k_0}}\Big) \\ &= \Omega\Big(\frac{1}{S} n^{2-\epsilon(n)}\Big) \end{aligned}$$

This completes the proof of Theorem 1.

References

[BC] A. Borodin and S. Cook, "A time-space tradeoff for sorting on a generl sequential model of computation," *SIAM Journal on Computing* **11** (1982), pp. 287-297.

[BFMUW] A. Borodin, F. Fich, F. Meyer auf der Heide, E. Upfal, and A. Wigderson, "A time-space tradeoff for element distinctness," *SIAM Journal on Computing* **16** (1987), pp. 97-99.

[BFKLT] A. Borodin, M. Fischer, D.Kirkpatrick, N. Lynch, and M. Tompa, "A time-space tradeoff for sorting on oblivious machines," *Journal of Computer and System Sciences* **22** (1981), pp. 351-364.

[C] A. Cobham, "The recognition problem for the set of perfect squares," Research Paper RC-1704, IBM Watson Research Center, Yorktown Heights, NY, April 1966.

[DG] P. Duris and Z. Galil, "A time-space tradeoff for language recognition," *Mathematical Systems Theory* **17** (1984), pp. 3-12.

[J] D. B. Johnson, "A simple proof of a time-space trade-off for sorting with linear comparisons," *Theoretical Computer Science* **43** (1986), pp. 345-350.

[K] M. Karchmer, "Two time-space tradeoffs for element distinctness," *Theoretical Computer Science* **47** (1986), pp. 237-246.

[Th] C. D. Thompson, "Area-time complexity for VLSI," *Proceedings Eleventh Annual ACM Symposium on the Theory of Computing*, May 1979, pp. 81-88.

[To] M. Tompa, "Time-space tradeoffs for computing functions using connectivity properties of their circuits," *Journal of Computer and System Sciences* **20** (1980), pp. 118-132.

[U] J. D. Ullman, *Computational Aspects of VLSI*, Computer Science Press, 1984.

[Y] A. C. Yao, "On the time-space tradeoff for sorting with linear queries," *Theoretical Computer Science* **19** (1982), pp. 203-218.